NOAA Technical Memorandum NWS SR-187

AN EVALUATION OF RIVER FORECAST MODEL OUTPUT FOR SIMULATIONS WITH AND WITHOUT QUANTITATIVE PRECIPITATION FORECASTS (QPF)

William B. Reed
Billy G. Olsen
John A. Schmidt

Arkansas-Red Basin RFC
Tulsa, Oklahoma

Scientific Services Division
Southern Region
Fort Worth, TX

May 1997

ITED STATES
DEPARTMENT OF COMMERCE
William M. Daley, Secretary

National Oceanic and Atmospheric Administration
James Baker
Under Secretary and Administrator

National Weather Service
Elbert W. Friday
Assistant Administrator

AN EVALUATION OF RIVER FORECAST MODEL OUTPUT FOR SIMULATIONS WITH AND WITHOUT QUANTITATIVE PRECIPITATION FORECASTS (QPF)

1. Introduction

This report provides results of a study which compares the observed river stages at five forecast points to corresponding forecast river stages. The forecasts were prepared with, and without, QPF input in order to evaluate the effect of the latter. The forecast river stages were derived by using the National Weather Service River Forecast System-Interactive Forecast Program (NWSRFS-IFP) which is used for forecast operations at the Arkansas-Red Basin River Forecast Center (ABRFC) in Tulsa, Oklahoma.

The ABRFC archive data files for the time period April 1 to June 11, 1995 were used to create a data set of NWSRFS-IFP river forecast model output for daily simulations with and without Quantitative Precipitation Forecasts (QPF). These dual simulations were made for five "headwater" basins: Chikaskia River above Corbin, Kansas (CBNK1); Whitewater River above Towanda, Kansas (TOWK1); East Cache Creek above Walters, Oklahoma (WLTO2); Illinois River above Watts, Oklahoma (WTTO2); and Spring River above Waco, Missouri (WCOM7). Table 1 summarizes basic headwater basin information.

Table 1. Headwater Basins

	CBNK1	TOWK1	WLTO2	WTTO2	WCOM7
Local Drainage Area (square miles)	794	426	675	635	739
Time to Unit Hydrograph Peak (hours)	6-12	18-30	42-54	12-24	18-24
Length of Unit Hydrograph (hours)	36	90	66	42	78
Flood Stage (ft)	10	22	21	13	19

The time period (April 1 to June 11, 1995) was selected because it was a period in which flooding occurred. The five basins were selected using two criteria: 1) the basin was to be a headwater basin and 2) the basin was to have flooded at least once during the selected time period. It was also a goal to select a basin in the respective Hydrologic Services Area (HSA) for each weather office that had been providing QPF to ABRFC since February 8, 1994, as part of the Southern Region Operational QPF Project (NWS 1994). These weather offices are Wichita, Dodge City, Tulsa, Amarillo, and Norman.

The simulations were made for only the 1200 UTC run time since this is when QPF information was available during the selected time period. The QPF consisted of four discrete 6-hr forecasts collectively covering the 24-hr period ending at 1200 UTC the following day. During floods, the ABRFC routinely runs the model at 1200 UTC, 1800 UTC, 0000 UTC, and 0600 UTC. However, this type of extended operations was not simulated.

This paper provides a description of the methodology used, the results of the data evaluation, a discussion of the results, and the resulting study conclusions.

2. Methodology

Archived NWSRFS and QPF files were used to create hydrologic forecasts with and without QPF. Dual 1200 UTC model runs were simulated for 72 days. After getting the model runs to fit the observed data up through 1200 UTC, the researcher recorded the forecast for 16 future times starting at 1800 UTC (each point being separated by 6 hr, i.e., each forecast is at the end of a 6-hr time step). A spreadsheet was created containing the two data sets as well as the 72 days of observed values for the five basins. These data were then evaluated by calculating the Root-Mean-Square-Error (RMSE) as described in Panofsky and Brier (1965). Additionally, these data were visually inspected for hits, misses, and false alarms.

For purpose of this study, a flood event comprises the period from when the river went above flood stage to when the river fell below. Using this definition, there were a total of 17 floods in the observed data. A hit is defined as when a river was forecast to go above flood stage and it did. A miss is defined as when the river was forecast to remain below flood stage when it went above. A false alarm is defined as when the river was forecast to go above flood stage and it did not. For this analysis of hits, misses, and false alarms, all forecasts were rounded to the nearest foot. The frequency of each type of event and the time step when each event occurred were noted.

For those who may want to conduct a similar study, a more complete description of the methodology used is provided in the Appendix.

3. Results

The results of the RMSE analysis for each of the five basins—calculated using 72 days of data—are presented in Figs. 1, 3, 5, 7, and 9. For comparison purposes the length of the unit hydrograph, flood stage, and the RMSE as a percent of flood stage at 96 hr are also provided in the respective basin figure. The combined RMSEs for the five basins are presented in Fig. 11. The number of hits, misses, and false alarms for both the "With QPF" and the "Without QPF" simulations are presented in Fig. 12. The occurrence of the hits, misses, and false alarms for the "With QPF" simulations are presented by time step in Fig. 13. The occurrence of the hits, misses, and false alarms for the "Without QPF" simulations are presented by time step in Fig. 14.

4. Discussion

The unit hydrograph is the means by which effective rainfall (derived from observed precipitation, QPF, or a combination of both) is converted to flow at the basin's outlet. Flow is then converted to stage by the use of a site-specific rating curve. For each basin, the hour at which the RMSE for "With QPF" diverges from the RMSE for "Without QPF" is a function of the "shape" of that basin's unit hydrograph and the magnitude and timing of the 6-hr QPF values. The hour at which the divergent series once again merge is a function of the same factors as well as the "length" of that basin's unit hydrograph. A visual inspection of Figs. 1, 3, 5, 7, and 9 and the corresponding hydrographs, Figs. 2, 4, 6, 8, and 10, reveals that the results are in good agreement with the above stated expected data properties. For example, consider basin CBNK1:

> Figure 2 shows runoff is minimal during the first period of effective rainfall. Therefore, as is depicted in Fig. 1, the simulations should not diverge during the first future time step, i.e., significant QPF-derived runoff from the basin is delayed in accordance with the shape of the unit hydrograph. It is important that the time series ***did not*** diverge quickly. If they had, the difference would not have been due to QPF. Once runoff begins, it lasts until the water has had time to leave the basin from the most distant upstream point, which corresponds to the end of the last step of the unit hydrograph. The unit hydrograph (Fig. 2) has a length of 36 hr (six 6-hr time steps). Therefore, if significant rainfall had been forecast during the fourth future time step, one would expect the time series to begin to converge at 54 hr (18 hr, the beginning of the significant rainfall + 36 hr, the length of the unit hydrograph = 54 hr). It is important that the time series ***did*** begin to converge at or before 54 hr. If they had not, the difference again would not have been due to QPF.[1]

On Figs. 1, 3, 5, 7, and 9, the RMSE at 6 hr is less than the RMSE at 96 hr. This is true for both simulations. This indicates that the uncertainty of RFC forecasts increases as the forecasts are extended farther into the future (especially during flood season). Also, the RMSE as a percentage of flood stage at 96 hr ranged from 23% to 41%. This may indicate that several of the RFC's basin models need to be recalibrated. However, these data properties did not mask the clear trend in the data sets. For all basins, the RMSEs of the "With QPF" data set ***were less than*** the RMSEs of the "Without QPF" data set.

The difference in the magnitude of the RMSE values from one station to the next is primarily a function of the difference in the rating curves for the stations. The differences in rating curves are caused by differences in channel capacities. For example, a rise of 3 ft would flood a 2-ft deep dry channel, but a rise of 3 ft would be insignificant if the dry channel were 30 ft deep. However, if both channels are already at bankfull, then a rise of 3 ft would cause flooding at both sites.

The RMSE values for the first four forecast periods calculated using 72 days of data and averaged for all five basins are presented in Table 2.

---

[1] It is interesting to note that Fig. 11 represents the range of hydrograph extremes, the RMSEs diverge quickly like TOWK1 (Fig. 3), and do not converge like WCOM7 (Fig. 9). The "bulge" displayed in the data (see also Figs. 1, 3, 5, 7, and 9) is caused by the general shape of the five unit hydrographs, i.e., they rise to a peak and then fall off. Therefore, the difference between models should increase to a maximum, then bulge, and then decrease.

Table 2. RMSE for First Four Forecast Periods (First 24 Hr)				
	1st 6-Hr Step	**2nd 6-Hr Step**	**3rd 6-Hr Step**	**4th 6-Hr Step**
Persistence	1.4 ft	2.2 ft	2.8 ft	3.4 ft
Without QPF	1.2 ft	2.1 ft	2.4 ft	3.2 ft
With QPF	1.0 ft	1.7 ft	1.8 ft	2.5 ft

The values for persistence were derived by using the observed 1200 UTC river stage at each forecast point as the forecast stage for the four forecast times (1800 UTC at 6 hr in the future, 0000 UTC at 12 hr in the future, 0600 UTC at 18 hr in the future, and 1200 UTC at 24 hr in the future). It can be seen that the RMSE values for persistence are greater than those for the model without QPF. Also, it can be seen that the RMSE values for the model without QPF are greater than those for the model with QPF. When making a forecast for 24 hr in the future, rather than just using persistence, a 26% improvement can be gained by using the model with QPF. In general, QPF improves model output at peak flow by approximately 20%.[2]

The RMSE for the model with QPF at 24 hr (2.5 ft) compared to an average flood stage of 17 ft indicates an overall model error of 15% at this forecast time. This lackluster performance can be attributed in part to the need for more forecaster interaction with both the model and the model input. When preparing the data sets for this study, the researchers used only the archived data without attempting further analysis of model input and without trying to simulate other data inquiries that take place during real-time forecasting. For example, often during real-time forecasting the river forecaster will interact with the RFC Hydrometeorological Analysis and Support (HAS) forecaster in a synergistic way. This interaction between river forecaster and HAS forecaster was not simulated.

The same number of hits and misses were forecast by the "With QPF" and the "Without QPF" simulations. However, this does not mean that the hits were forecast with the same accuracy or with the same lead time. The RMSE analysis showed that the "With QPF" simulations were more accurate. A comparison of Figs. 13 and 14 shows that the "With QPF" simulations provided more lead time for one event, i.e., a hit that occurred in the first time step without QPF (Fig. 14) occurred in the fifth time step with QPF (Fig. 13)—this provided 24 hr more lead time.[3]

With regard to misses, note that all the study forecasts were model simulations, not the actual real-time forecasts made by a forecaster. Even under this constraint, it is likely that a simulation

[2] This was estimated by using the difference between the model RMSEs at hour 48 on Fig. 11 divided by the larger error [(5.12 - 4.08) / 5.12 = 0.20]. The average improvement using all 16 forecast times is 15%. The range is 2.4% at hour 96 to 20.4% at hour 42. It is interesting to note that Fig. 11 shows improvement at all 16 forecast times.

[3] After observing the difference between Figs. 13 and 14, the spreadsheet was reviewed to verify that the difference was caused by the same event.

of operations under which QPF is provided twice daily (as it now is, at 1200 and 0000 UTC), has the potential to greatly reduce the number of simulated misses in both the "With QPF" and the "Without QPF" simulations. Additionally, many of the storms that cause flooding within ABRFC's area occur overnight. Therefore, a study that included the 0000 UTC model run would likely result in fewer misses for the "With QPF" simulations than for the "Without QPF" simulations. During real-time operations, a miss is unlikely because the model is run routinely three times per day and more often when there is a threat of flooding.

It would have been unusual to have a false alarm without QPF, since the precipitation used in the model for the "Without QPF" simulations has already occurred. For there to be a false alarm during the "Without QPF" simulations:

- the model would have to be poorly calibrated for the basin(s), or
- the data quality control for the simulation(s) would have to be absent, or
- the initial carryover values for variable states used in the simulation(s) would have to be in error.

Therefore, ***it is not surprising*** that the false alarms occurred only in the "With QPF" data set. One false alarm occurred in the fourth time step, and the other three occurred in the fifth time step (Fig. 13). Issuing a watch instead of a warning—when a flood is forecast based only on QPF and does not occur within the next 24 hr—has the potential to reduce false alarms by 75% (Fig. 8). Using no QPF eliminates false alarms, but this would be at the sacrifice of accuracy and at the sacrifice of better lead times. In summary, QPF is beneficial to river forecasting, especially if a watch is issued instead of a warning for those events when a flood is forecast based only on QPF and does not occur within the next 24 hr.

5. Conclusions

- The results of this study show that the use of QPF adds value to the river forecasting process. QPF improves model output at peak flow by approximately 20%.
- The RMSE values show that the model, regardless of whether QPF is used, is reliable (i.e., the model outperforms persistence). However, without additional forecaster interaction, the model, although reliable, is subject to errors. Additionally, without QPF the model will most likely underforecast river stages until after the effective precipitation has ended.
- The noise and trend in the values of the RMSE suggest the models need to be run several times per day during the events. (The models are routinely run three times per day at ABRFC and are run more often during flood events.)
- Since QPF adds value to hydrologic forecasts, as with the river forecast model, QPF needs to be updated more often than once per day during events. Indeed, it can be inferred

from Zipser (1990) that to take full advantage of satellite and radar observations, the precipitation forecasts should be updated every 12 hr, if not more frequently. (ABRFC now routinely receives QPF at 1200 and 0000 UTC.)

- To avoid false alarms (Fig. 13), perhaps flood warnings beyond 24 hr (beyond the first four 6-hr periods) should not be issued if they are based only on QPF. This seems to be a reasonable tradeoff between providing adequate advanced warning and reducing the potential for false alarms.

- This report supports the argument that weather offices should consider issuing a river flood watch rather than a river flood warning for those events when the flood is based only on QPF *and* the flood is not forecast to occur within the next 24 hr.

- This report also supports the argument that RFCs should consider issuing forecasts that provide the weather offices with both the "With QPF" and the "Without QPF" time series. The "With QPF" time series would be the official forecast. The "Without QPF" time series would be for informational purposes only. This change in RFC operations is suggested for seven reasons:

    - It supports issuance of watches by the weather office (improves hydrologic services to the public).

    - It supports operations at weather offices (the product from the RFC has to be understood and interpreted by the weather office before being issued to public).

    - It supports basin-wide, river-forecast QPF verification at the RFC by seamlessly integrating operations and verification in a manner that can be implemented at ABRFC with only minor changes to existing operations. (A prototype program has already been written which uses as its input the proposed product format that includes both with and without QPF river forecasts. The program output is similar to Figs. 1, 3, 5, 7, and 9 of this paper).

    - It supports training at the weather office (evaluation of basin response to QPF).

    - It improves operations at the RFC (evaluation of basin/model response to QPF).

    - It aids coordination between weather offices and RFCs during events.

    - It reduces the potential for false alarms by supporting the issuance of watches by the weather offices.

ACKNOWLEDGMENTS

The authors would like to thank Steve Amburn and Steve Hrebenach at NWSFO Tulsa, and Ed May and Dan Smith at Southern Region Headquarters, for their reviews of this manuscript in draft form and helpful suggestions.

REFERENCES

Panofsky, H. A. and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. Pennsylvania State University, University Park, PA, 224 pp.

National Weather Service, 1994: Southern Region Operational QPF Project. NWS Southern Region Technical Attachment SR/HSD 94-7. Fort Worth, TX. 3 pp.

Zipser, E. J., 1990: Rainfall Predictability: When Will Extrapolation-Based Algorithms Fail? Preprints, *8th AMS Conference on Hydrometeorology*, (Alberta, Canada), Amer. Meteor. Soc., Boston, MA.

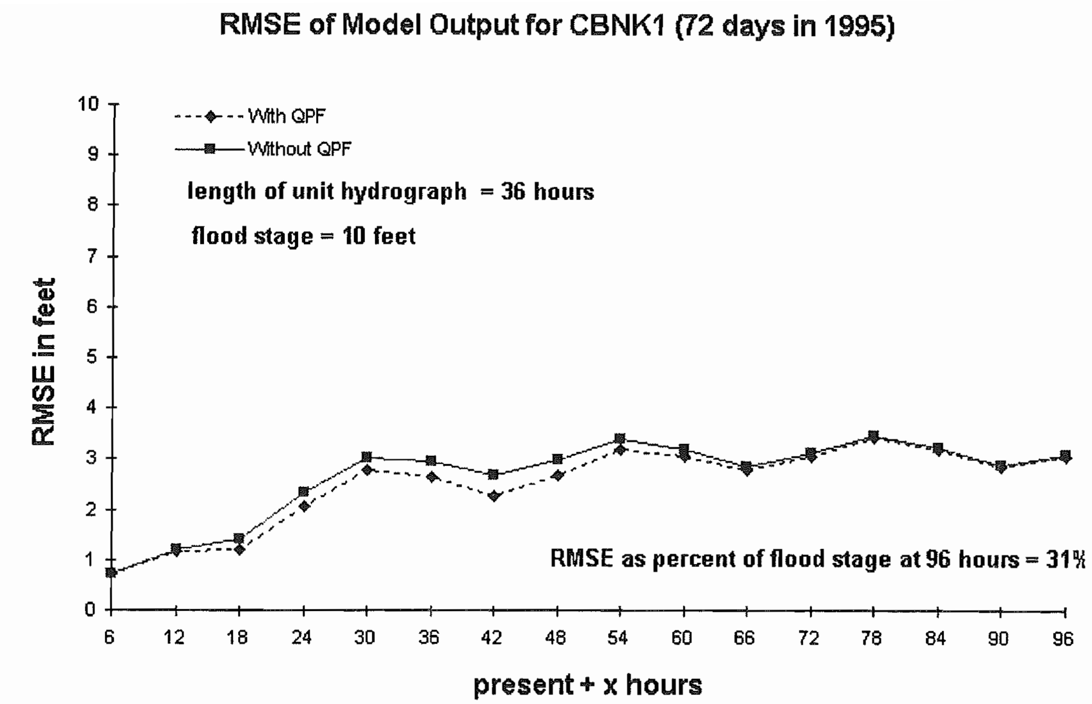

Figure 1.

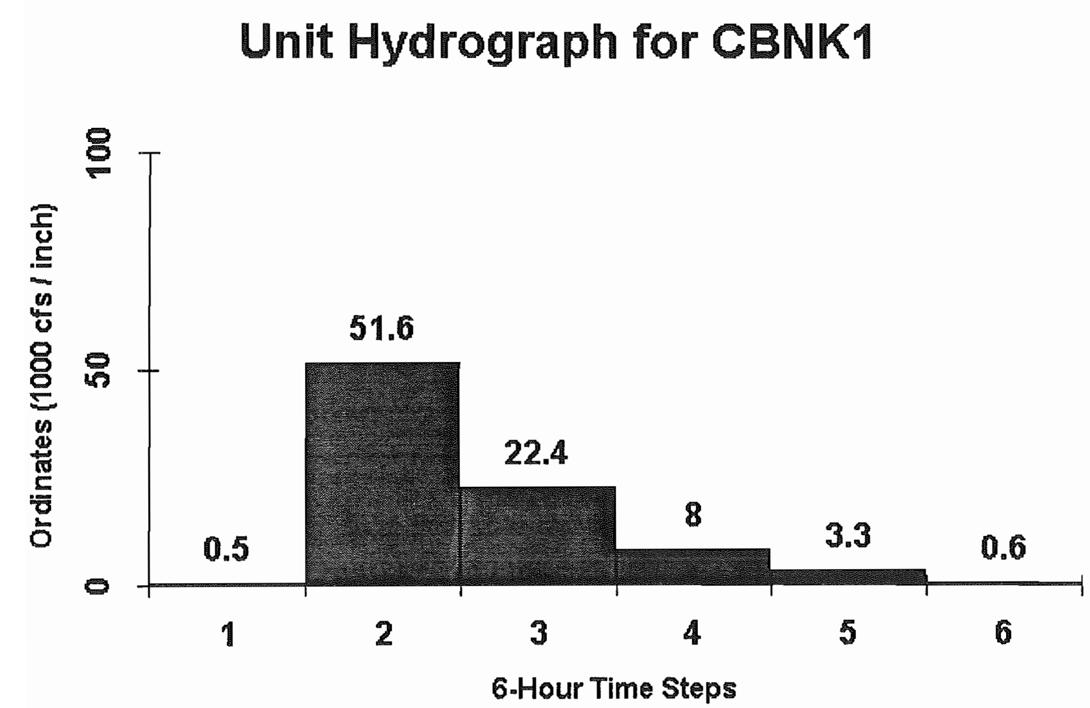

Figure 2.

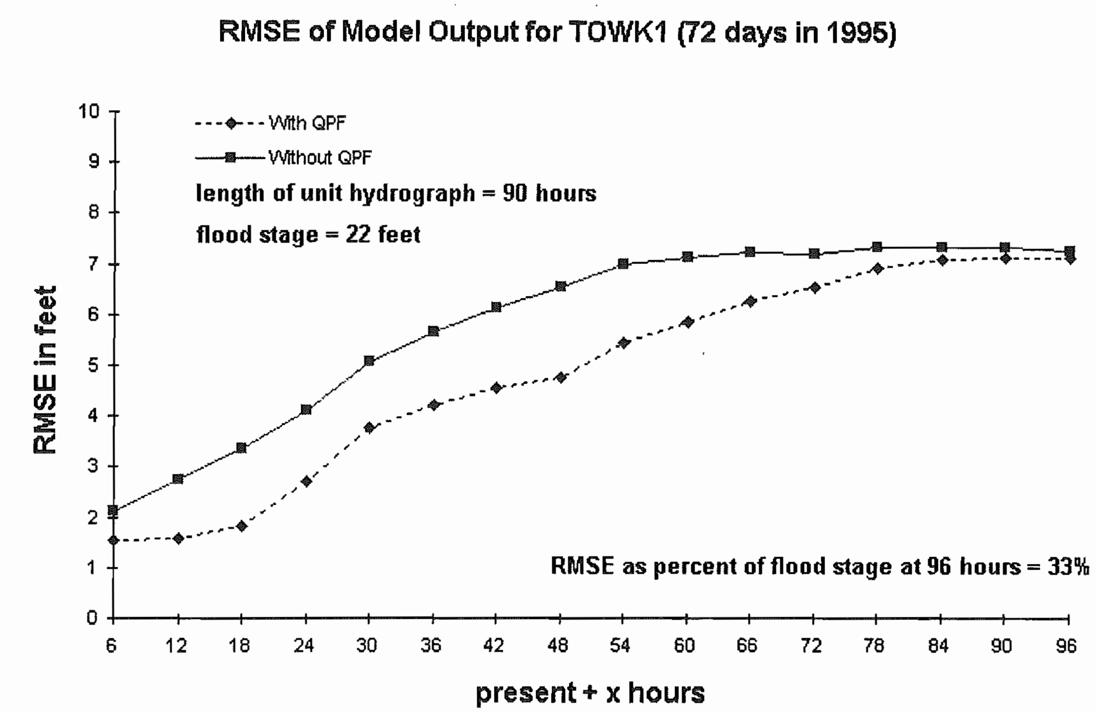

Figure 3.

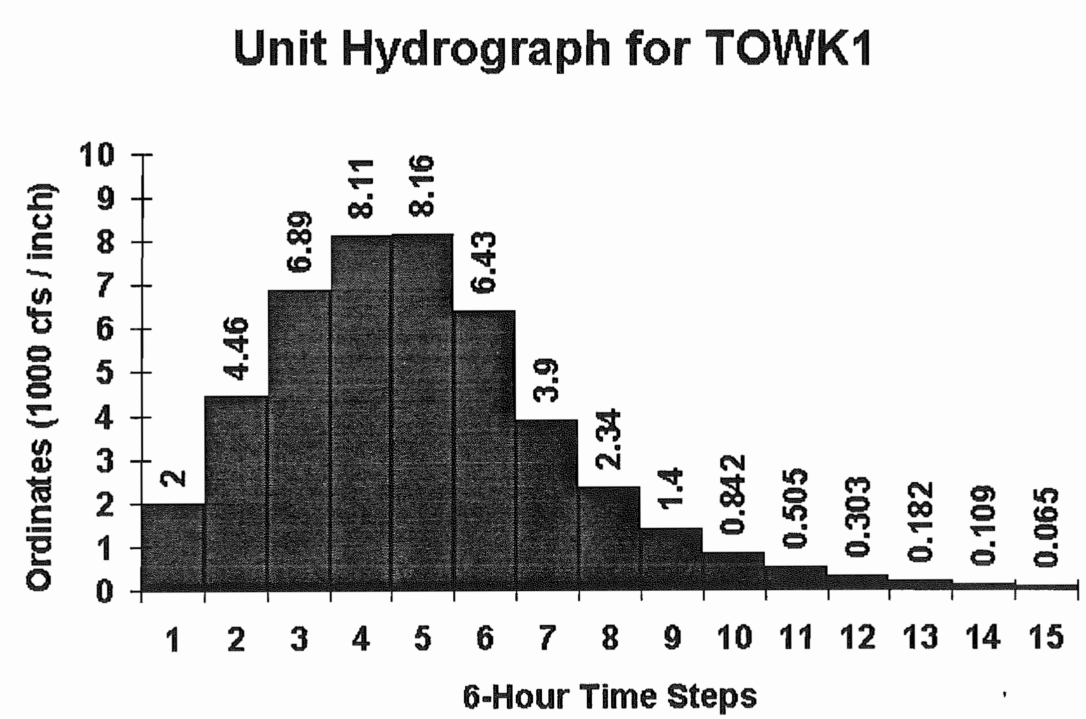

Figure 4.

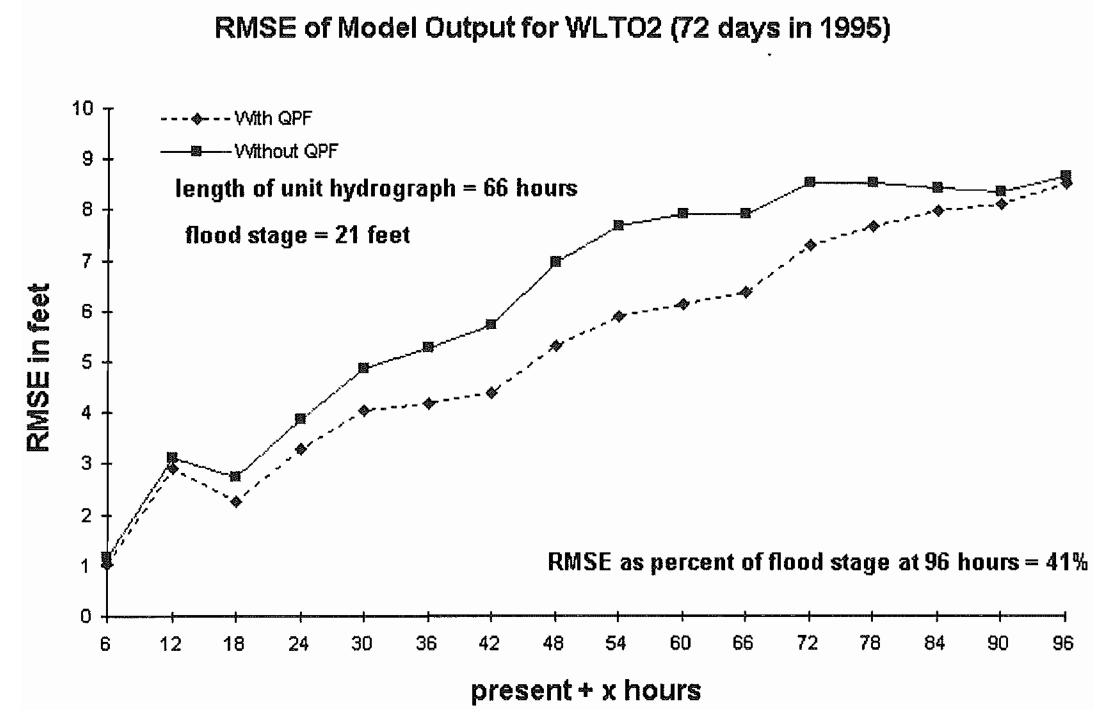

Figure 5.

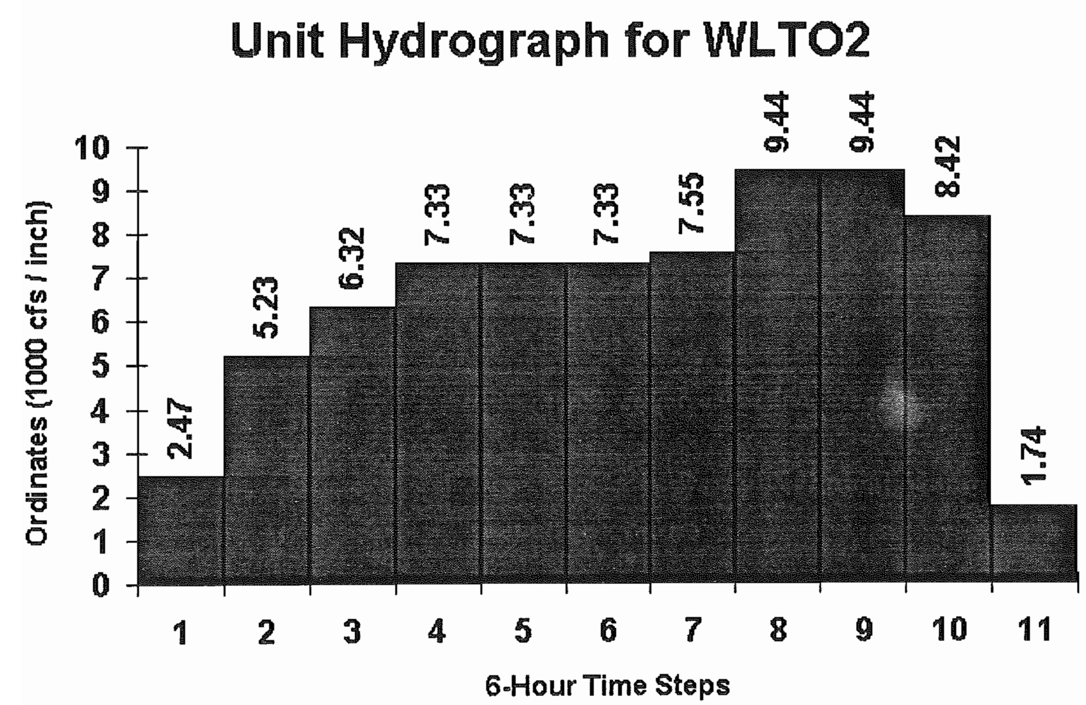

Figure 6.

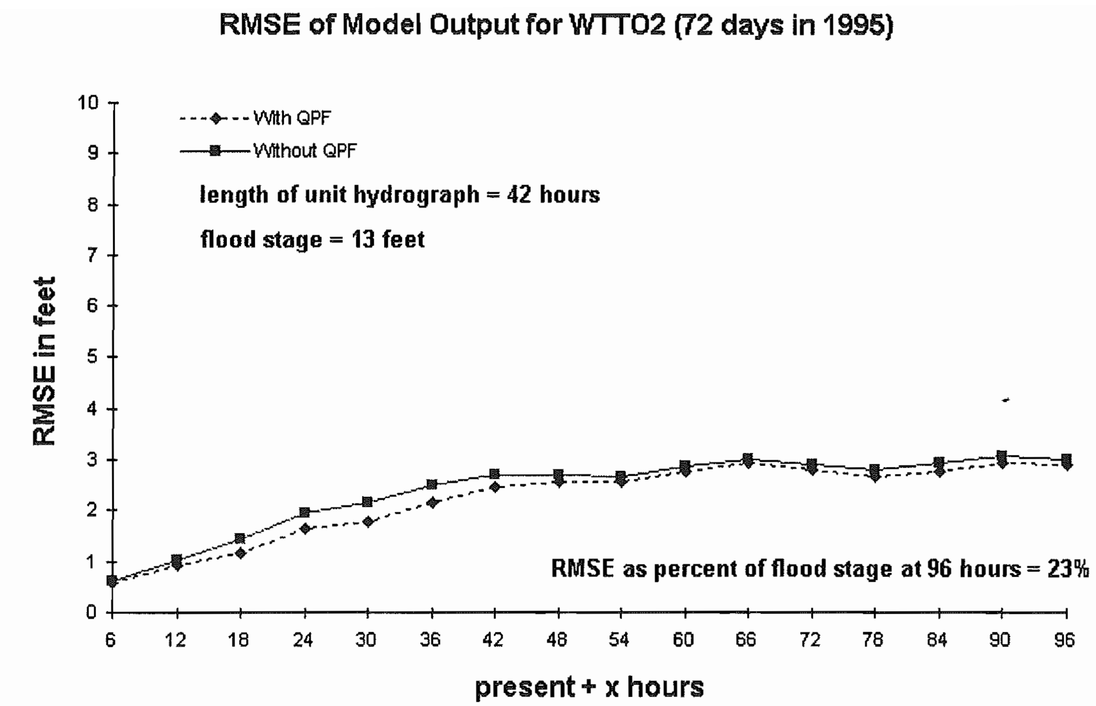

Figure 7.

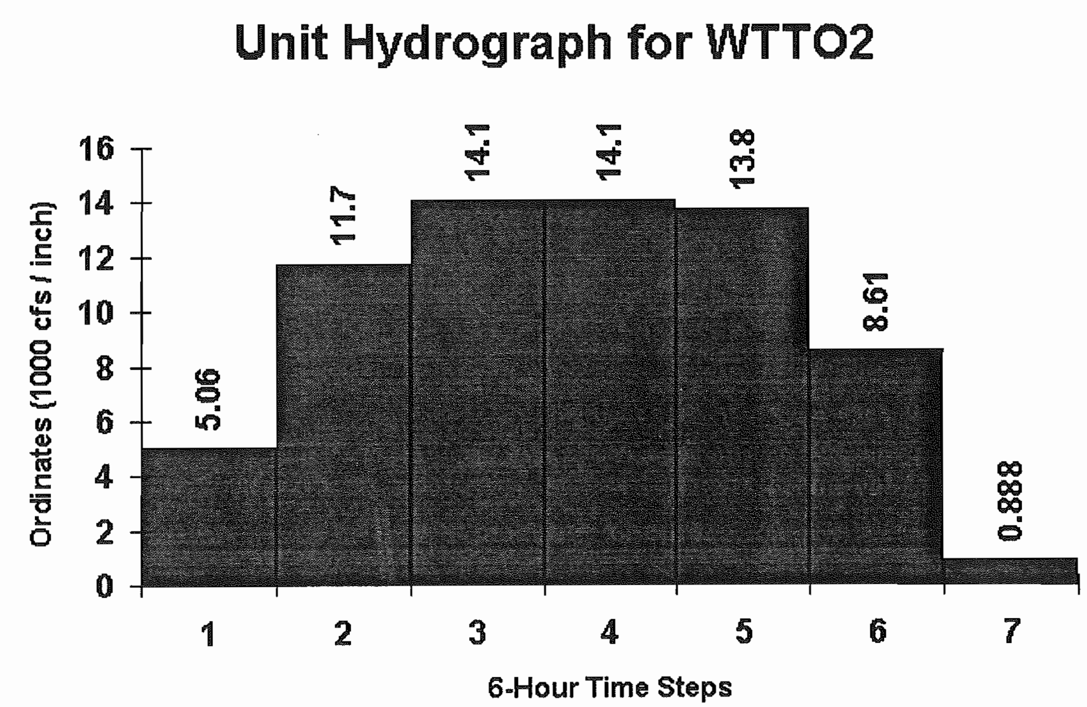

Figure 8.

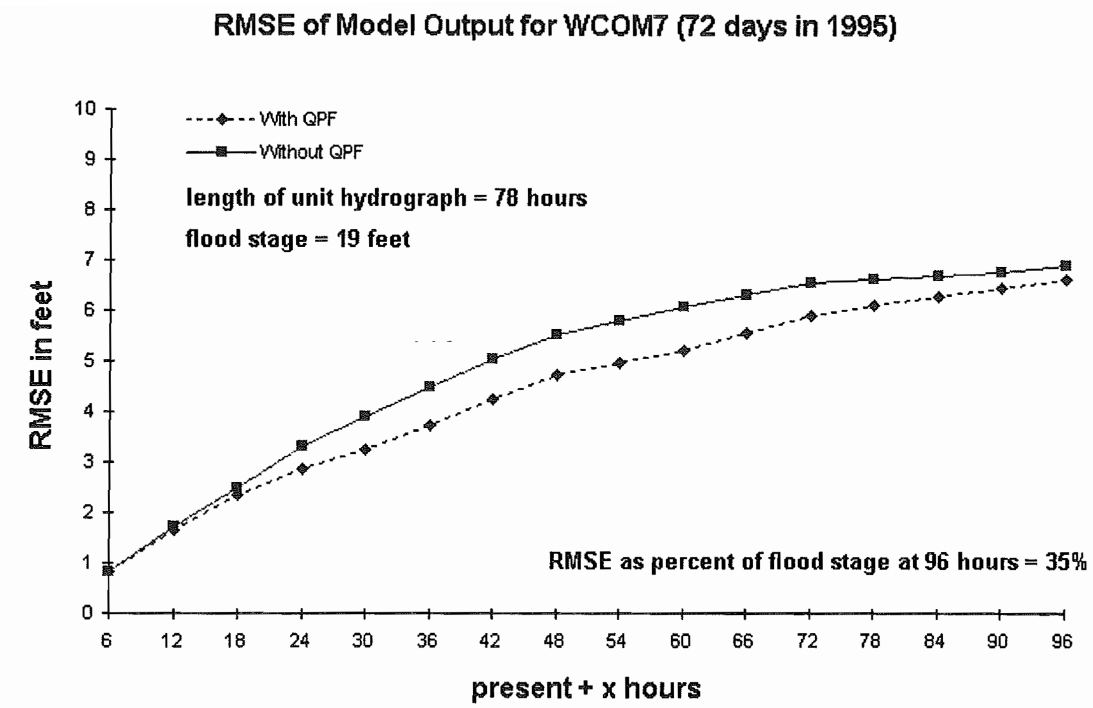

Figure 9.

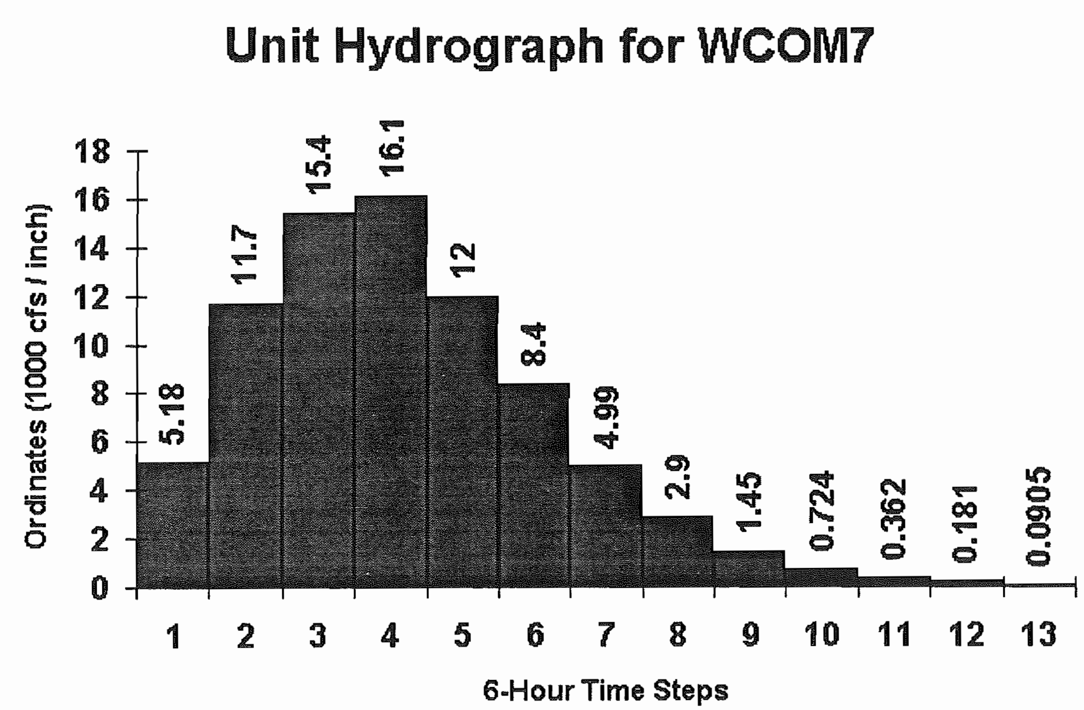

Figure 10.

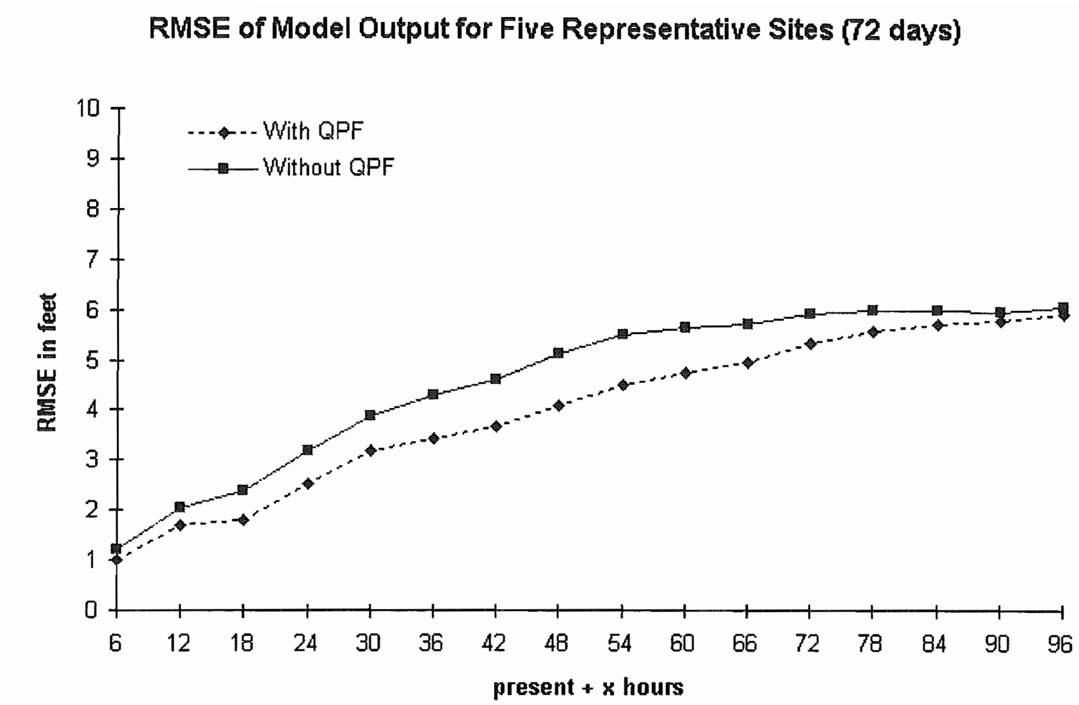

Figure 11.

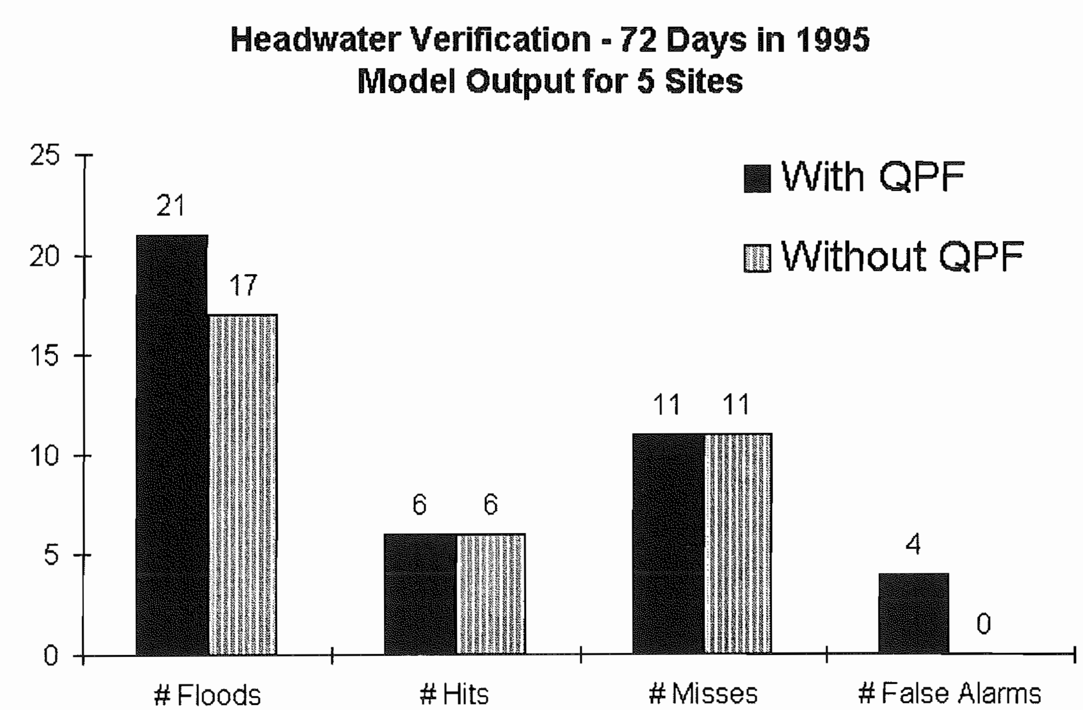

Figure 12.

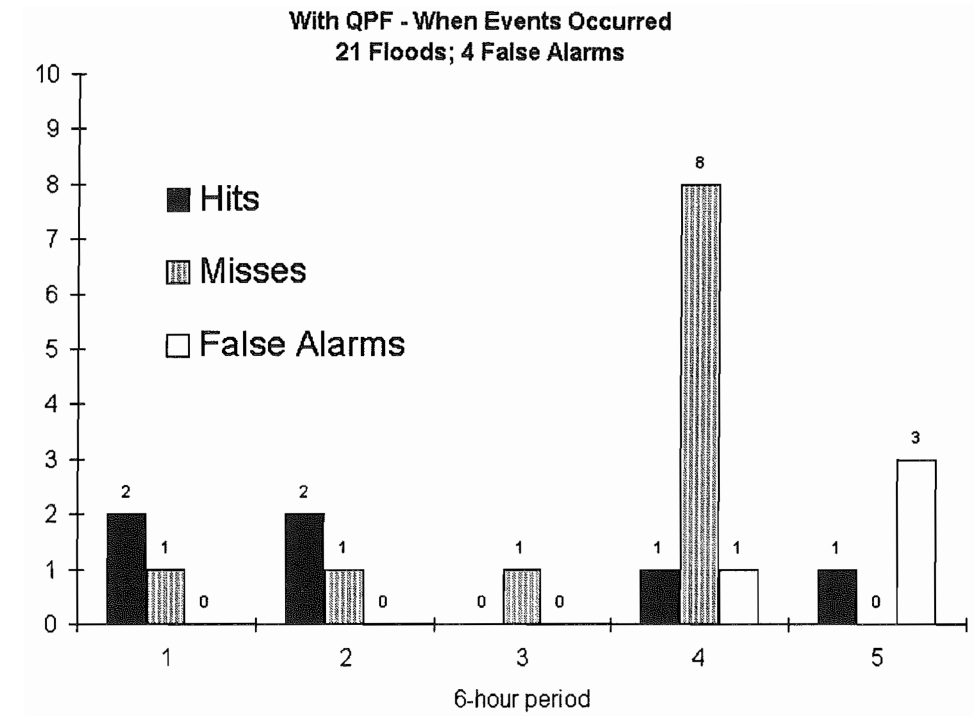

Figure 13.

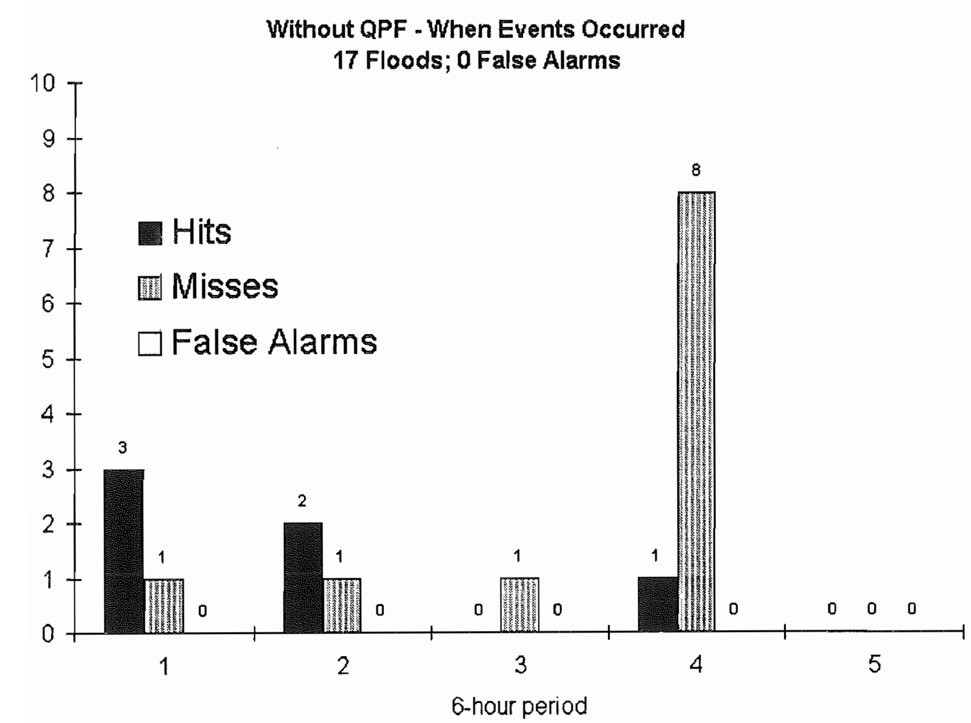

Figure 14.

Appendix

DETAILED DESCRIPTION OF THE METHODOLOGY USED FOR THE HEADWATER VERIFICATION STUDY

Archived NWSRFS Operational Forecast System (OFS) fs5files files (the operational set of data and parameter files created by NWSRFS-OFS needed to run NWSRFS-IFP) were retrieved from CD-ROMs prepared by the RFC on a routine basis. Also retrieved were the QPF future mean areal precipitation (fmap) files for the same time period. The fmap file is the output file created by the software the RFC uses to mosaic the QPF input files received from the QPF support offices. The format of the fmap file is a Standard Hydrologic Exchange Format (SHEF) encoded time series file. The archive OFS files were copied as needed for individual dates to a special directory. This directory was then used as the specified ofs_level and ofs_inpt_grp in the researcher's local .Apps_defaults. (The .Apps_defaults is a resource file where NWSRFS looks for application defaults. Within .Apps_defaults ofs_level and ofs_inpt_grp are tokens pointing to the directory where IFP looks for the OFS files needed.) Thus when the researcher started IFP, an IFP script copied the retrieved archive files from the special directory to the directory used by IFP. The researcher then selected the appropriate forecast group, carryover date, run date, run time, and forecast point. The run time was set to 1200 UTC for all simulations. All future precipitation beyond 1200 UTC was turned off by setting the universal technique switch for future precipitation in IFP to zero. This was an important step because otherwise the historical observed values for rainfall beyond 1200 UTC would have been used by the model.

After using run-time modifications (mods) to fit the observed data up through 1200 UTC the researcher then recorded the forecast for 16 future times starting at 1800 UTC (each point being separated by 6 hours). These forecasts were 6 to 96 hours in the future from the set time and date. The next step was to then simulate the same run date and run time with QPF. This was accomplished by using the basin specific values from the retrieved fmap file for the run date. The values were then added to the simulation as a Rainfall/Runoff Input Change (RRICHNG) mod. The RRICHNG mod is used to change the amount of moisture entering a rainfall/runoff model for a specific basin. Both mean areal precipitation (map) values based on observations and fmap values based on QPF can be entered. The segment was then rerun with this mod and the researcher recorded the "new " forecasts at the same 16 future times.

This dual process was accomplished for five sites and 72 days were simulated Therefore, 720 simulations were made resulting in 11,520 data points (5,760 data points in the "Without QPF" data set and 5,760 data points in the "With QPF" data set).

To allow a fair comparison between data sets, the Change Times Series (TSCHNG) mod was not utilized during simulations. The TSCHNG mod is used to overwrite the adjusted instantaneous river discharge (QINE) time series of model calculated flow values. Also, the mods between the dual simulations were kept the same except for the addition of the RRICHNG mod to the "With QPF" simulation.

All the basins simulated were headwater basins except for WCOM7. Since WCOM7 is not a headwater basin, the simulations for this site required that the upstream basin routed to WCOM7 be run with and without QPF respectively, prior to the corresponding WCOM7 simulation. This upstream basin is the headwater basin for Spring River above Carthage, Missouri (CTHM7). Although simulations were made for CTHM7, the resulting data were not used in this study. However, the resulting data for WCOM7 were used. Additional information:

- WCOM7 total drainage area = 1,164 square miles
- WCOM7 local drainage area = 739 square miles
- upstream of WCOM7 is CTHM7 (Spring River above Carthage, Missouri)
  - lag from CTHM7 to WCOM7 = 9 hours,
  - length of CTHM7 unit hydrograph = 60 hours,
  - CTHM7 drainage area = 425 square miles

A spreadsheet was created containing the two data sets as well as the 72 days of observed values for the five basins. These data were then evaluated by calculating the Root-Mean-Square-Error (RMSE) as described in Panofsky and Brier, 1968. For each of the five basins, the RMSE was calculated for each of the 16 forecasts per day (6-96 hours in the future). Additionally, these data were visually inspected for hits, misses, and false alarms . For purpose of this study a flood event comprises the period from when the river went above flood stage to when the river fell below. Using this definition, there were a total of 17 floods in the observed data. A hit is defined as when a river was forecast to go above flood stage and it did. A miss is defined as when the river was forecast to remain below flood stage when it went above. A false alarm is defined as when the river was forecast to go above flood stage and it did not. For this analysis of hits, misses, and false alarms, all forecasts were rounded to the nearest foot. The frequency of each type of event, and the time step when each event occurred were noted.